

## Exploring RNA Structural Codes with SHAPE Chemistry

KEVIN M. WEEKS\* AND DAVID M. MAUGER

*Department of Chemistry, University of North Carolina Chapel Hill, North Carolina 27599-3290, United States*

RECEIVED ON FEBRUARY 25, 2011

### CONSPECTUS

**R**NA is the central conduit for gene expression. This role depends on an ability to encode information at two levels: in its linear sequence and in the complex structures RNA can form by folding back on itself. Understanding the global structure–function interrelationships mediated by RNA remains a great challenge in molecular and structural biology. In this Account, we discuss evolving work in our laboratory focused on creating facile, generic, quantitative, accurate, and highly informative approaches for understanding RNA structure in biologically important environments.

The core innovation derives from our discovery that the nucleophilic reactivity of the ribose 2'-hydroxyl in RNA is gated by local nucleotide flexibility. The 2'-hydroxyl is reactive at conformationally flexible positions but is unreactive at nucleotides constrained by base pairing. Sites of modification in RNA can be detected efficiently either using primer extension or by protection from exoribonucleolytic degradation. This technology is now called SHAPE, for selective 2'-hydroxyl acylation analyzed by primer extension (or protection from exoribonuclease).

SHAPE reactivities are largely independent of nucleotide identity but correlate closely with model-free measurements of molecular order. The simple SHAPE reaction is thus a robust, nucleotide-resolution, biophysical measurement of RNA structure. SHAPE can be used to provide an experimental correction to RNA folding algorithms and, in favorable cases, yield kilobase-scale secondary structure predictions with high accuracies.

SHAPE chemistry is based on very simple reactive carbonyl centers that can be varied to yield slow- and fast-reacting reagents. Differential SHAPE reactivities can be used to detect specific RNA positions with slow local nucleotide dynamics. These positions, which are often in the C2'-endo conformation, have the potential to function as molecular timers that regulate RNA folding and function. In addition, fast-reacting SHAPE reagents can be used to visualize RNA structural biogenesis and RNA–protein assembly reactions in one second snapshots in very straightforward experiments.

The application of SHAPE to challenging problems in biology has revealed surprises in well-studied systems. New regions have been identified that are likely to have critical functional roles on the basis of their high levels of RNA structure. For example, SHAPE analysis of large RNAs, such as authentic viral RNA genomes, suggests that RNA structure organizes regulatory motifs and regulates splicing, protein folding, genome recombination, and ribonucleoprotein assembly. SHAPE has also revealed limitations to the hierarchical model for RNA folding. Continued development and application of SHAPE technologies will advance our understanding of the many ways in which the genetic code is expressed through the underlying structure of RNA.



### Introduction

The central role of RNA in biology reflects its unique abilities to encode genetic information in its linear sequence and to fold into structures that contribute to diverse biological functions. RNA structures may selectively present or conceal access to the information stored in a linear RNA sequence. Alternatively,

the structures themselves may be important; these structures can be as simple as the small hairpins processed to become miRNAs or as sophisticated as ligand binding switches, complex protein binding elements, and catalytic active sites.

The structure of an individual RNA is determined by the complex pattern of interactions among nucleotide bases at

both secondary and tertiary structure levels. The secondary structure consists of base-pairing interactions that form helices and define loops in individual structural elements. Tertiary interactions are longer-range and bring individual structural elements together to create higher-order structures. There are strong hints that transcription initiation codes,<sup>1</sup> splicing codes,<sup>2,4</sup> translation initiation codes,<sup>3</sup> and protein folding codes<sup>4,5</sup> are all expressed in the form of RNA structure. Thus, the raw genetic information encoded in the primary sequence of RNA is communicated by additional information encoded in the RNA secondary and tertiary structure.

We have only begun to glimpse these RNA codes because the structures of the vast majority of potential regulatory and functional elements in RNA are unknown. The current situation resembles playing on a complex jungle gym with one's eyes closed. Experience in our laboratory emphasizes the importance of the following principles. First, each RNA has many possible secondary structures, but only a few are highly populated. Second, biologically important differences between structures can be subtle. Thus, identifying populated states of an RNA requires structural information for every nucleotide. Third, short fragments of long RNAs often do not fully recapitulate the structure that exists in the authentic RNA. Moreover, folding is subject to very strong "end effects" such that either omitting an outlying element or adding flanking sequences may induce misfolding. Thus, full-length, biologically active RNAs should be studied. Fourth, the fine details and dynamics that contribute to RNA folding are important. Thus, sequence specificities, time dependencies, and other idiosyncrasies of experiments designed to probe the RNA structure must be well understood. In this Account, we discuss the applications of selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE), a collection of technologies capable of measuring local nucleotide flexibility at single-nucleotide resolution for RNAs of arbitrary size in diverse biological environments.<sup>6</sup>

### Simple, Generic, Covalent Chemistry at the RNA 2'-Hydroxyl

Our initial vision was to find a simple, generic approach to probe simultaneously the structure of every nucleotide within an RNA.<sup>6</sup> Early work from our laboratory showed that chemical reactivity at the 2'-ribose position (Figure 1A) is exquisitely sensitive to the local nucleotide environment.<sup>7-10</sup> We discovered that a wide variety of electrophilic reagents, especially anhydrides and acyl cyanides, react at the 2'-hydroxyl group, and that reactivities are strongly gated

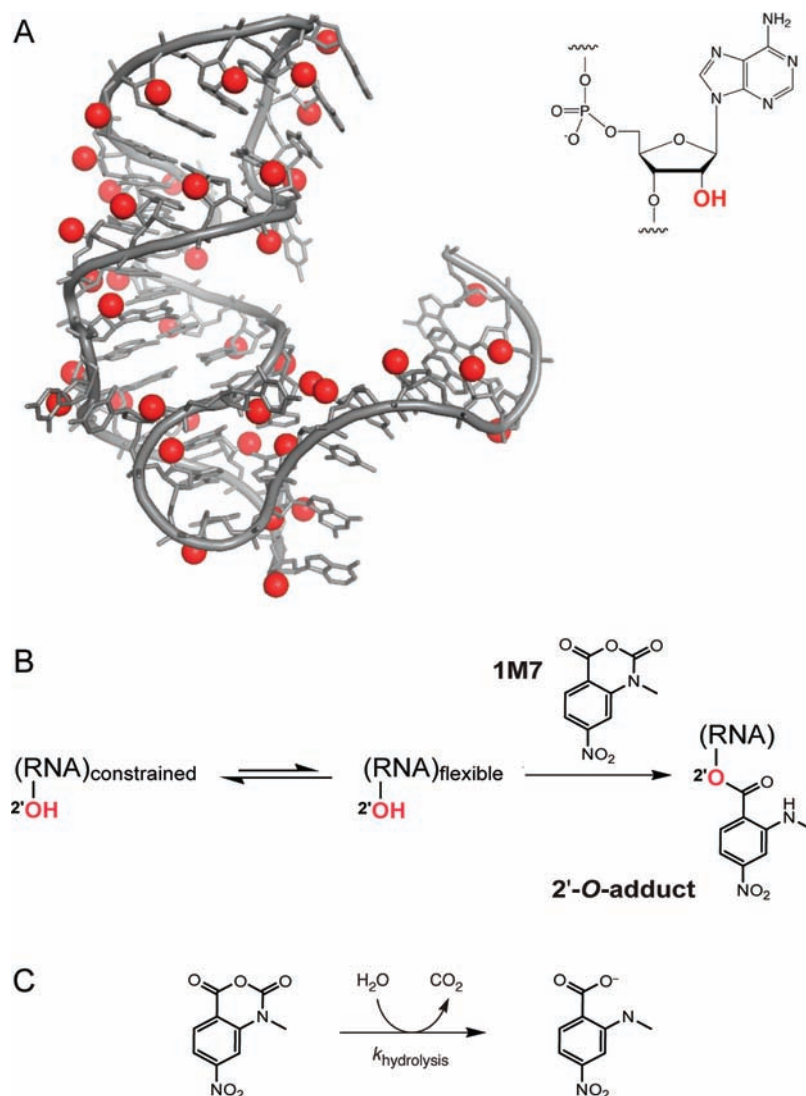
by local nucleotide flexibility (Figure 1B). Flexible nucleotides sample multiple conformations, a few of which strongly facilitate nucleophilic attack by the 2'-hydroxyl. The reactivity of individual nucleotides is determined by electrostatic communication between the 2'-OH and adjacent 3'-phosphate diester<sup>6,9</sup> and other features intrinsic to favorable ribose conformations (McGinnis and Weeks, unpublished results).

Electrophilic reagents capable of reacting with RNA 2'-hydroxyl groups also undergo simultaneous inactivation by hydrolysis with water (Figure 1C). This competing hydrolysis reaction imparts three advantages to SHAPE experiments. First, the hydrolysis reaction makes RNA probing extremely straightforward to perform. The hydroxyl-selective electrophile either reacts with RNA or is inactivated by hydrolysis: No explicit quench step is required. Second, although reaction with RNA is the experimental focus of SHAPE, the solution lifetime of a given SHAPE reagent is determined primarily by reagent hydrolysis with ~55 M water. Thus, SHAPE modification of RNA is insensitive to both the RNA concentration and complications due to additional cross-reacting components, including buffer components, small molecules, or proteins. Third, as will be outlined below, hydrolysis rates for distinct SHAPE reagents vary over a large range and by comparing reactivities of different reagents it is possible to examine RNA folding kinetics and dynamics in very straightforward experiments.

### The Basic SHAPE Experiment

RNA structure analysis involves treatment of an RNA with a SHAPE reagent, such as isatoic anhydride (IA), *N*-methyl isatoic anhydride (NMIA), 1-methyl-7-nitro-isatoic anhydride (1M7), or benzoyl cyanide (BzCN) (Scheme 1).<sup>6,11,12</sup> Reaction results in the selective formation of covalent adducts at the 2'-ribose position of flexible nucleotides. Reaction conditions are controlled to yield sparse modification of the RNA, typically modification of 1 in 300 nucleotides. The detected signal thus reflects uncorrelated adduct-forming events on unmodified RNA or in unmodified regions of a long RNA. The reaction is allowed to proceed until the reagent is consumed by either reaction with RNA or hydrolysis with water. For the general analysis of RNA structure, the 1M7 SHAPE reagent<sup>11</sup> (Figure 1B) is easy to handle, reacts in ~1 min, and reports RNA secondary and tertiary interactions accurately.

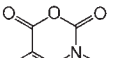
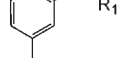
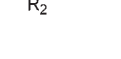
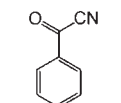
Following selective 2'-hydroxyl acylation (Figure 2A), covalent SHAPE adducts can be detected by reverse transcriptase-mediated primer extension.<sup>6,13</sup> DNA synthesis by



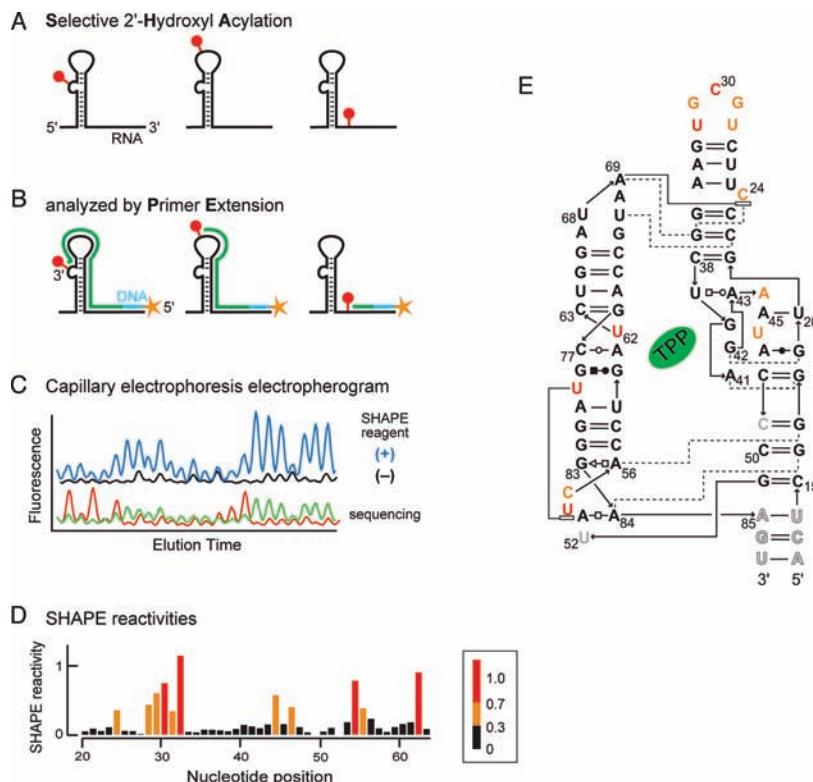
**FIGURE 1.** The RNA 2'-hydroxyl group and SHAPE chemistry: (A) position of the 2'-hydroxyl group; (B) mechanism of the SHAPE reaction to form 2'-O-adducts at flexible nucleotides; (C) mechanism of the parallel hydrolysis reaction.

reverse transcriptase stops one nucleotide prior to the position of an adduct (Figure 2B).<sup>6,10</sup> The length of each cDNA reports the site of a SHAPE modification in the original RNA. A control primer extension reaction is performed in parallel using unmodified RNA to locate sites of natural reverse transcriptase pausing and pre-existing RNA degradation. Peaks in both the modified and no-reagent control lanes are assigned by comparison with sequencing ladders (Figure 2C). Subtraction of the intensity of modified RNA peaks from intensities of no-reagent control peaks yields a reactivity profile (Figure 2D).<sup>13,14</sup> Nucleotides constrained by base-pairing and tertiary interactions have low SHAPE reactivities whereas single-stranded and unconstrained nucleotides have higher reactivities, as illustrated for the TPP riboswitch aptamer domain (Figure 2E).

**SCHEME 1.** A Toolbox of Useful SHAPE Reagents

| Reagent   | R <sub>1</sub>  | R <sub>2</sub>  | hydrolysis<br>$t_{1/2}$ (s) |   |
|---|-----------------|-----------------|-----------------------------|---|
|  <b>IA</b>   | H               | H               | 430                         | <div style="display: flex; align-items: center;"> <div style="border-left: 1px solid black; height: 100%; margin-right: 5px;"></div> <div style="writing-mode: vertical-rl; transform: rotate(180deg);">slow</div> </div> |
|  <b>NMIA</b> | CH <sub>3</sub> | H               | 260                         |   |
|  <b>1M7</b>  | CH <sub>3</sub> | NO <sub>2</sub> | 14                          |   |
|  <b>BzCN</b> | –               | –               | 0.25                        |   |

High-throughput SHAPE can be performed using fluorescently labeled primers and capillary electrophoresis using a



**FIGURE 2.** Overview of the SHAPE Experiment: (A) RNA is selectively modified (red dots) at flexible nucleotides in an RNA. (B) Positions of adduct formation are detected by primer extension. (C) Primer extension products from the experimental, no-reagent control, and sequencing markers are resolved by capillary electrophoresis. (D) Electropherograms are computationally deconvoluted to yield normalized SHAPE reactivities (see scale). (E) Superposition of SHAPE reactivities on a secondary structure model for the TPP riboswitch aptamer domain. Adapted from ref 16. Copyright 2010 American Chemical Society.

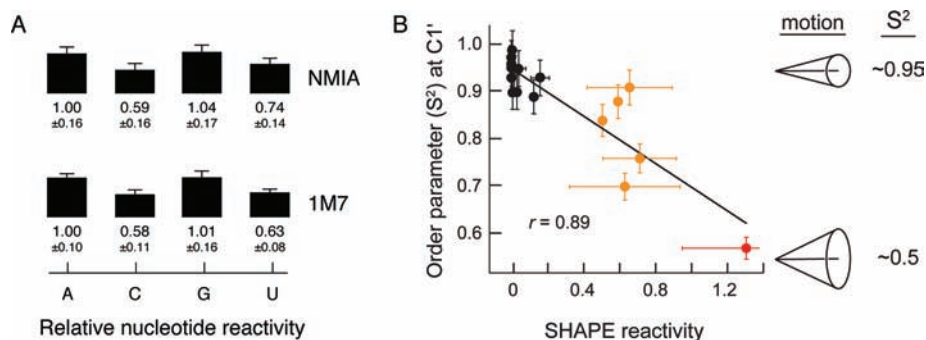
commercial DNA sequencing instrument<sup>15</sup> and analyzed using custom software.<sup>14</sup> A single high-throughput SHAPE experiment typically provides data on several hundred RNA nucleotides and is particularly well suited for analyzing the structure of large RNAs.<sup>15</sup> 2'-*O*-Adducts can also be detected by their ability to protect an RNA from degradation by a 3' → 5' exoribonuclease (RNase-detected SHAPE).<sup>16</sup> RNase-detected SHAPE allows analysis of short RNAs and nucleotides very close to the ends of an RNA. SHAPE will eventually be melded with next-generation sequencing readouts.

### SHAPE as a Quantitative Measure of Nucleotide Flexibility

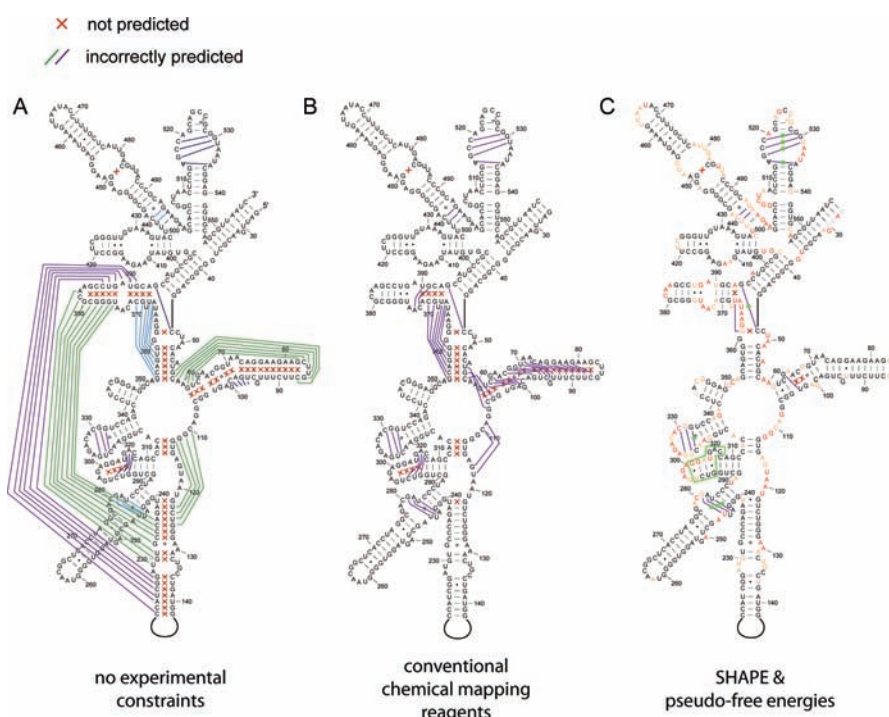
Our laboratory has pursued three classes of experiments to place SHAPE chemistry on a rigorous biophysical foundation. First, analysis of nucleotide reactivity preferences for NMIA and 1M7 shows that adenosine and guanine react approximately 1.6-fold more rapidly than do cytosine and uridine (Figure 3A). These small, but statistically significant, differences appear to have no effect on the ability of SHAPE to measure quantitative conformational features at all

nucleotides in an RNA. Variations in nucleotide reactivity for SHAPE reagents are much smaller than the biases found for other chemical probes that form covalent adducts with RNA.<sup>17</sup>

Second, nucleotide-resolution dynamics in RNA can be measured by NMR using the model-free generalized order parameter,  $S^2$ ,<sup>18,19</sup> which varies from 0.0 (disordered) to 1.0 (fully ordered). Nucleotides that are highly ordered have large  $S^2$  values (~0.95), whereas nucleotides that sample a larger conformational space have much lower  $S^2$  values (~0.5).<sup>19</sup> SHAPE and NMR analyses were performed on three well-characterized RNAs.<sup>20</sup> Nucleotide resolution SHAPE reactivities and the NMR-derived  $S^2$  parameter correlate strongly. For example, for the RNA regulatory element that binds the U1A protein,  $r$  is 0.89 (Figure 3B). This correlation is impressive given that  $S^2$  measurements are derived primarily from NMR measurements on the picosecond to nanosecond time scale, whereas SHAPE is likely influenced by these rapid motions and by motions on a slower time scale. From these experiments, we conclude that SHAPE measures spatial disorder at single nucleotide resolution.



**FIGURE 3.** SHAPE reactivities are independent of nucleotide identity and correlate quantitatively with local nucleotide flexibility. (A) Intrinsic nucleotide-specific reactivities for NMIA and 1M7.<sup>17</sup> (B) Correlation between the model-free generalized order parameter,  $S^2$ , and SHAPE reactivity in the U1A-binding RNA element. Low, medium, and high SHAPE reactivities are emphasized in black, yellow, and red, respectively. Adapted from refs 17 and 20.

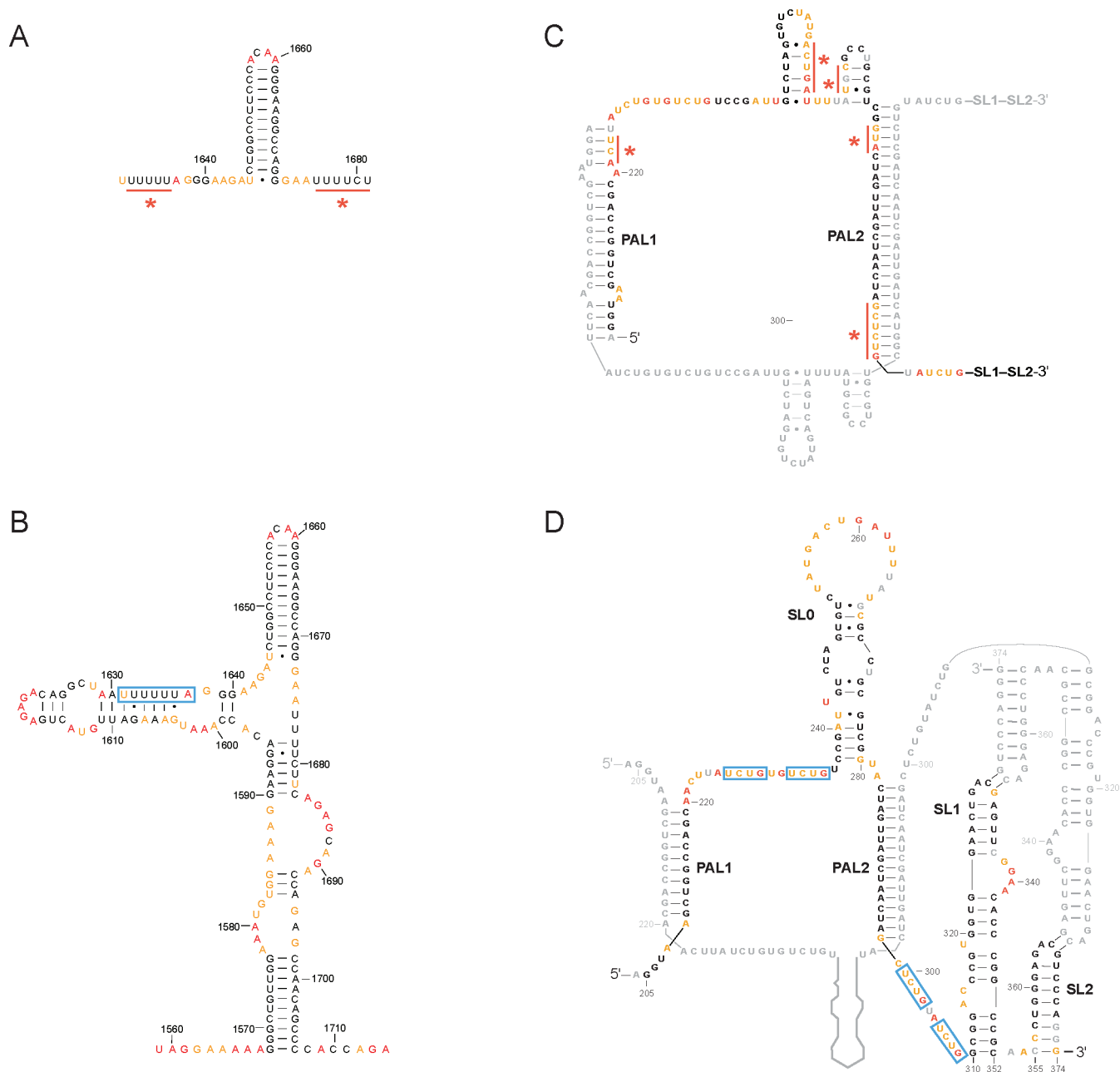


**FIGURE 4.** Comparison of RNA secondary structure predictions for the 5' domain of *E. coli* 16S rRNA, using (A) no experimental constraints, (B) conventional chemical probing reagents, or (C) SHAPE reactivity-based pseudo-free-energy change terms ( $\Delta G_{\text{SHAPE}}$ ) as constraints.<sup>27</sup> Base pairs in the accepted structure that are missing in the prediction (×) and base pairs that are predicted incorrectly (green and blue lines) are indicated. Green boxes and dots indicate areas in the covariation structure where SHAPE reactivities support formation of alternate base pairs. Adapted from ref 27.

Third, it was formally possible that SHAPE reactivity might be modulated by the physical accessibility of the 2'-hydroxyl group. Although this group is solvent accessible in an isolated helix, it is often buried in folded RNA structures (Figure 1A). Multiple experiments on simple duplexes and on RNAs with complex tertiary structures have shown that there is essentially no relationship between SHAPE reactivities and solvent accessibility.<sup>4,6,20–22</sup> Critically, many fully buried, but unconstrained, nucleotides react readily with SHAPE reagents. SHAPE thus probes the local nucleotide flexibility of RNA nucleotides regardless of solvent accessibility.

### Accurate SHAPE-Directed RNA Secondary Structure Prediction

Most computational methods predict RNA secondary structure by approximating the free energy of structural motifs as the sum of individual base-pairing energies within these motifs (reviewed in refs 23 and 24). This works well for simple structures, especially hairpins, but is not accurate for large RNAs. We lack thermodynamic parameters for all possible sequence elements, do not understand the contributions of higher-order interactions to free energy, and cannot predict the influences of RNA folding kinetics or interactions with

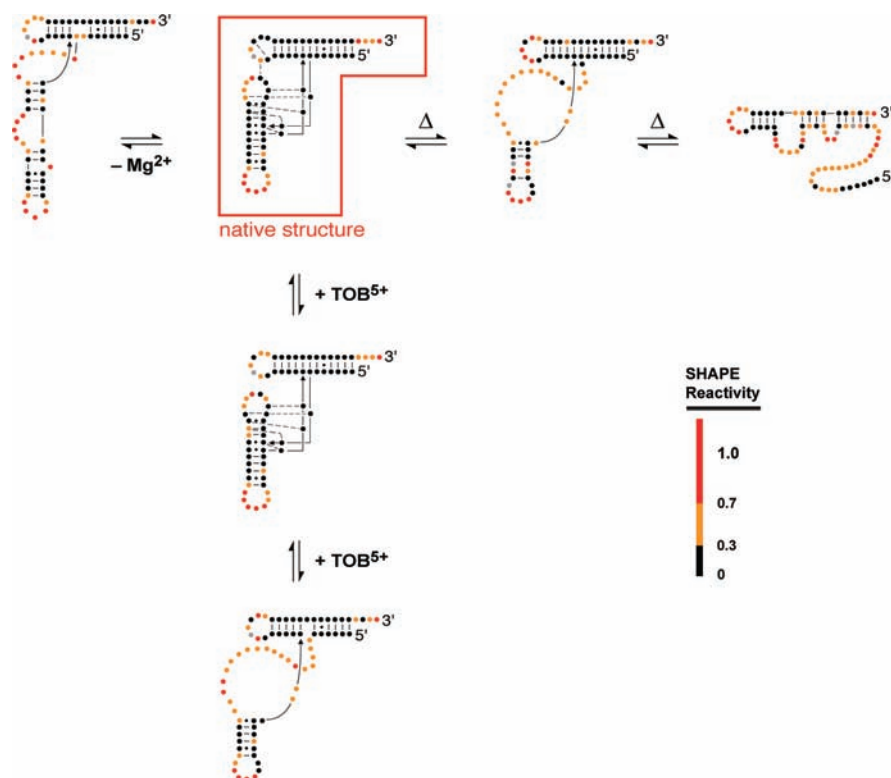


**FIGURE 5.** SHAPE reactivities support revised RNA secondary structure models. (A) Original model for the HIV-1 frameshift element<sup>31</sup> and (B) revised model based on SHAPE data.<sup>4</sup> (C) Initial model for the viral dimerization and packaging domain of MuLV<sup>34</sup> and (D) SHAPE-supported model.<sup>37</sup> Red bars and asterisks highlight areas where the SHAPE data are not consistent with a secondary structure model. Notable functional elements in the SHAPE-supported structures are highlighted by blue boxes.

proteins or small molecules. For example, when the secondary structure of the *Escherichia coli* 16S rRNA is predicted based on sequence alone using one of the most accurate programs currently available, RNAstructure,<sup>25,26</sup> only 50% of accepted base pairs are predicted correctly (Figure 4A).<sup>27</sup> Accuracies can be improved by incorporation of experimental information into the prediction algorithm.<sup>28</sup> In general, data obtained from small chemical probes improves

prediction accuracy to a much greater extent than data generated by RNase enzymes.<sup>29,30</sup> Prediction of 16S rRNA structure improves to 72% upon incorporation of data from traditional chemical probing reagents (Figure 4B).<sup>27</sup> However, this structure still contains serious errors.

As outlined above, SHAPE yields a true biophysical measurement of RNA dynamics. We thus converted SHAPE reactivities into a pseudo-free-energy change term



**FIGURE 6.** Analysis of complex RNA structural transitions by SHAPE. Model for structural changes in tRNA<sup>Asp</sup> induced by varying the Mg<sup>2+</sup> concentration, upon heating ( $\Delta$ ), or upon binding by tobramycin are shown.<sup>39,40</sup> Figure reproduced from ref 40. Copyright 2008 American Chemical Society.

( $\Delta G_{\text{SHAPE}}$ ) and used this as an experimental correction to an RNA folding algorithm. Using SHAPE pseudo-free-energies derived from reaction with the 1M7 reagent, the structures of 16S rRNA (Figure 4C) and a set of smaller RNAs ranging in size from 75 to 155 nt are predicted with accuracies of  $\sim 95\%$ .<sup>27</sup> In a few areas (Figure 4C), SHAPE analysis suggests that the 16S rRNA may adopt alternative conformations not found in the accepted model. In general, SHAPE-directed RNA secondary structure prediction has proven highly accurate; additional innovations will be required to predict structures in which only a few nucleotides are unconstrained and thus reactive.<sup>11,27</sup>

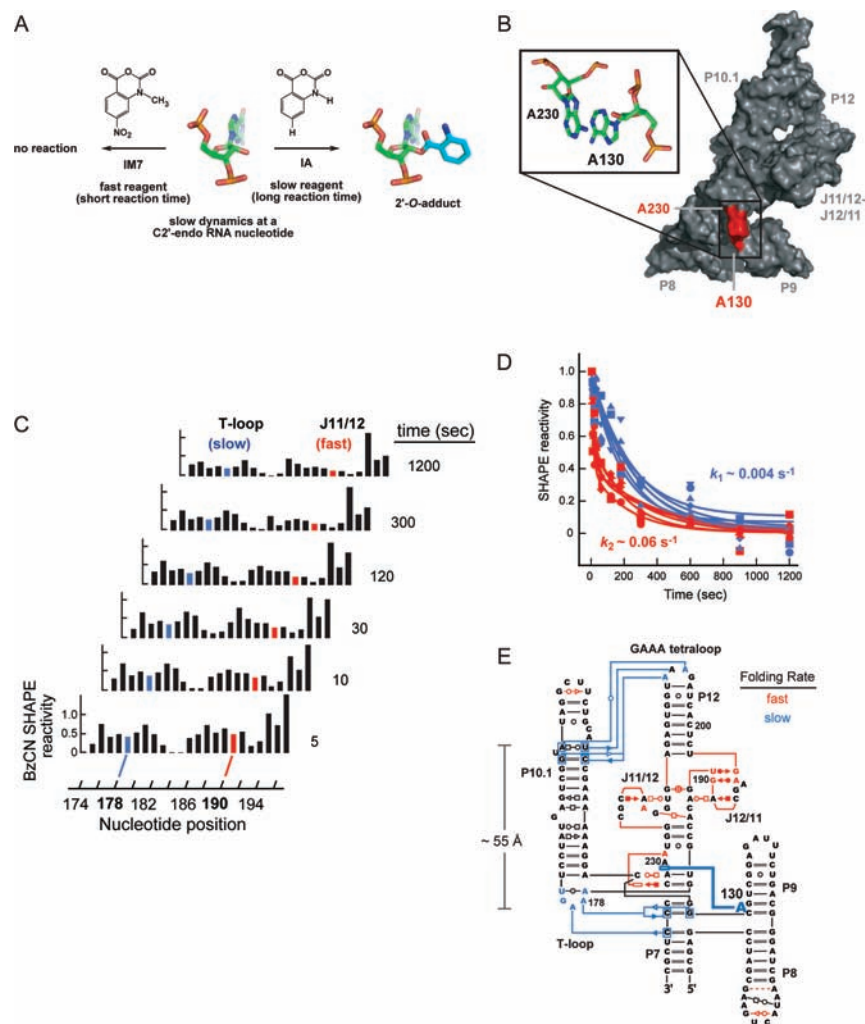
## Structural Surprises

The nucleotide-resolution data produced by SHAPE has consistently yielded structural surprises. Before showing examples in which SHAPE supports alternative models of secondary structure and folding, we do want to emphasize that SHAPE cannot prove that any of the proposed RNA structures are correct. However, any proposed model must be substantially compatible with SHAPE analyses.

The frameshift element in the HIV-1 RNA genome causes a  $-1$  frameshift to enable the translation of the HIV-1 *Po*

gene.<sup>31</sup> This element has traditionally been modeled as a simple stem–loop downstream of a single-stranded, poly-U slippery sequence (Figure 5A). SHAPE analysis of an intact, full-length HIV-1 genomic RNA revealed that several regions proposed to be single-stranded actually had low SHAPE reactivities (Figure 5A, red bars).<sup>4</sup> SHAPE-derived pseudo-free-energies suggest that the slippery sequence base pairs to an upstream sequence and the complete regulatory signal forms a large branched RNA domain (Figure 5B).<sup>4</sup> The revised model is consistent with previous findings that the HIV-1 frameshift element required additional sequences downstream of the simple stem–loop for full activity.<sup>32,33</sup> Phylogenetic analysis revealed that the structure proposed based on SHAPE data is conserved among divergent HIV-1 isolates.<sup>4</sup>

In a second example, the RNA genome of murine leukemia virus (MuLV) forms a dimer during a critical step in virion packaging. The dimer structure is minimally composed of two palindromic sequences (PAL1 and PAL2) that form intermolecular duplexes and two stem loops (SL1 and SL2) that form loop–loop interactions (reviewed in refs 34 and 35). SHAPE analysis of the MuLV RNA domain using authentic genomic RNA revealed several regions in which



**FIGURE 7.** Measuring RNA dynamics by SHAPE. (A) Use of fast (IM7) and slow (IA) reagents to detect nucleotides with slow conformational dynamics.<sup>43</sup> (B) Structural context of A130, which exhibits slow dynamics and functions as a molecular timer for folding of an RNase P specificity domain.<sup>44</sup> (C) Time-resolved SHAPE reactivities and (D) representative time–progress curves. (E) Model for time-resolved tertiary folding of the *B. subtilis* RNase P specificity domain.<sup>12</sup> Nucleotides that fold in the fast versus slow phase are emphasized in red and blue, respectively, in panels C–E. Adapted from ref 44.

nucleotides proposed to base pair actually had high SHAPE reactivities (Figure 5C).<sup>35–37</sup> The SHAPE-directed model of the dimerization domain suggests that the PAL1 and PAL2 duplexes are significantly shorter than traditionally drawn<sup>34</sup> and that many of the nucleotides between PAL1 and PAL2 form a large, bulged stem–loop structure, SLO (Figure 5D).<sup>37</sup> Subsequent work showed that the two sets of UCUG sequences, which are exposed and flanked by helices in the SHAPE-directed model but partially obscured in the prior model, constitute the viral packaging signal.<sup>38</sup>

In a third example, we monitored the unfolding of tRNA<sup>Asp</sup> transcripts as a function of three distinct perturbations: reduced Mg<sup>2+</sup> concentration, increased temperature, and addition of the cationic antibiotic tobramycin.<sup>39,40</sup> Although presumed to be simple, the SHAPE data revealed

that the energy landscape for the folding of this molecule contains many structurally diverse states: each of these three conditions denatured the tRNA in a unique way and yielded distinct end structures (Figure 6). These studies reveal that the folding of even a small RNA molecule can be very complex and nonhierarchical.

These three examples illustrate principles that apply broadly to the analysis of RNA structure using SHAPE. First, many important elements of prior models remain key features in the SHAPE-directed analysis, for example, the 1660 hairpin in HIV-1,<sup>31</sup> the UCUG elements in the dimerization domain of MuLV,<sup>41</sup> and the sensitivity of tRNA<sup>Asp</sup> structure to ligand binding. Second, SHAPE reveals structural details that may result in revision of accepted models. The revised structural models are substantially consistent with



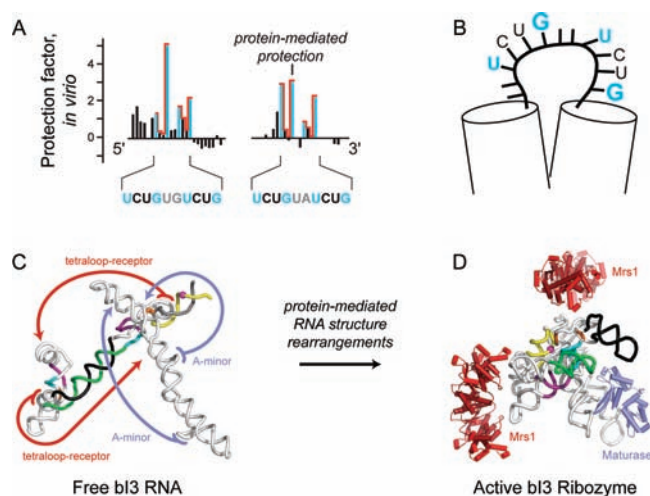
prior functional data, but these data are interpreted in a different way. Third, elements within SHAPE-directed RNA structure models sometimes do not correlate perfectly with the SHAPE data. This reflects, in part, that multiple structures can be formed at equilibrium and that a single model may be an oversimplification.<sup>37,40,42</sup> In addition, pseudoknots and other tertiary interactions need to be better accounted for in our algorithms.

## Analysis of RNA Dynamics Using SHAPE

In aqueous solution, SHAPE reagents react with RNA and also undergo inactivation by hydrolysis (Figure 1C). The rate of reaction with RNA is proportional to the rate of hydrolysis. Our laboratory has created a tool kit of SHAPE reagents (Scheme 1) that makes it possible to study the local folding dynamics of individual RNA nucleotides. There are two general approaches for using SHAPE to obtain time-resolved information regarding RNA folding: by differential reactivity and by using fast-reacting reagents.

When differential reactivities of slow- versus fast-reacting SHAPE reagents, such as IA versus 1M7 (Scheme 1) are compared, most nucleotides have identical reactivities.<sup>43,44</sup> However, a critical subset is reactive toward the slow reagent but unreactive toward the fast reagent (Figure 7A). These nucleotides exhibit very slow local nucleotide dynamics with half-lives on the order of minutes and many adopt the C2'-endo conformation. Nucleotides with a C2'-endo conformation are much less prevalent than C3'-endo nucleotides (characteristic of A-form RNA helices) but are highly over-represented in regions of RNA molecules with complex tertiary structures.<sup>44,45</sup> One such nucleotide is A130 in the specificity domain of the *Bacillus subtilis* RNase P ribozyme. A130 shows strong differential SHAPE reactivity, adopts the C2'-endo conformation, and mediates a critical tertiary interaction by stacking with nucleotide A230 (Figure 7B). These two nucleotides play a critical role in recognition of the tRNA substrate.<sup>46,47</sup>

The time frame of many RNA folding events is on the order of a few seconds to minutes. Benzyl cyanide (BzCN, Scheme 1) reacts completely in 1 s and is used in time-resolved SHAPE<sup>12,48</sup> to take snapshots of RNA structure (Figure 7C). Time-resolved SHAPE showed that the tertiary folding of the RNase P domain proceeds in two distinct, kinetically significant steps (Figure 7D).<sup>12</sup> Local tertiary structures involving the J11/12 and J12/11 loops fold at  $\sim 0.06 \text{ s}^{-1}$ . In contrast, interdomain tertiary interactions, like that involving A130 which spans 55 Å, fold much more slowly ( $\sim 0.004 \text{ s}^{-1}$ ; Figure 7E). Strikingly, deleting A130



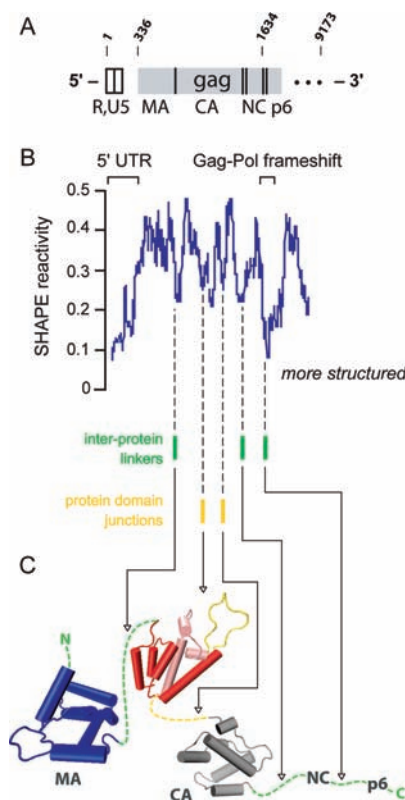
**FIGURE 8.** SHAPE analysis of RNA–protein interactions. (A) Binding of nucleocapsid to the MuLV genomic RNA results in protection of a small number of nucleotides from SHAPE modification (emphasized with red line). (B) Model for the nucleocapsid RNA binding site. Nucleotides most strongly protected are emphasized in blue; flanking helices are shown as cylinders.<sup>38</sup> (C) The bl3 group I intron RNA exists in a mis- and unfolded state prior to protein binding.<sup>52</sup> (D) Binding by the maturase and Mrs1 proteins stabilizes long-range tertiary interactions (blue and red arrows) and induces widespread structural rearrangements as shown by changes in SHAPE reactivities.<sup>21,49</sup> Adapted from refs 38 and 49.

sped up folding of the whole domain by an order of magnitude.<sup>44</sup> Together, differential and time-resolved SHAPE support the general model that certain C2'-endo nucleotides function as molecular timers to control RNA folding kinetics.

## RNA–Protein Interactions

SHAPE is a robust tool for analyzing RNA structure in biologically important RNA–protein complexes (RNPs). If a protein binds to a simple, single-stranded element, SHAPE can identify the precise binding site directly. Conversely, protein binding can promote large-scale structural rearrangement and thereby change SHAPE reactivity in many regions of an RNA.

As an example of detection of direct binding, we studied packaging of the retrovirus MuLV. Retroviruses have evolved conserved pathways to recognize genomic RNA dimers and to specifically exclude monomeric genomes and cellular RNAs. Packaging of RNA genomes into nascent virions is mediated by interactions with the nucleocapsid domain of the Gag protein. SHAPE reagents readily penetrate viral particles and were used to detect differences between protein-bound RNA inside virions and protein-free RNA that had been gently extracted from virions.<sup>15,38</sup> In addition, protein–RNA interactions were probed inside the



**FIGURE 9.** SHAPE reveals a relationship between RNA structure and protein domain structure in the HIV-1 genomic RNA. (A) Organization of the HIV-1 genome, emphasizing the Gag coding region. (B) Median SHAPE reactivities over a moving 75 nucleotide window. (C) Relationship between structured RNA regions and the peptide loops that link independent protein domains in Gag.<sup>4</sup>

viral particle in reverse-footprinting experiments in which nucleocapsid–RNA interactions were weakened by treatment with a reagent that disrupted the fold of the nucleocapsid domain. SHAPE analysis indicated that the nucleocapsid domain of Gag binds and constrains the first and last nucleotides within a tandemly repeated UUCG motif (in blue, Figure 8A), flanked by base-paired regions (Figure 8B).<sup>38</sup> This work demonstrates the ability of SHAPE to precisely detect protein–RNA interactions at single-nucleotide resolution, to reveal the overall architecture of a binding site, and to do so in a complex virus particle environment. Analogous studies identified key components of the packaging signal for HIV-1.<sup>15</sup>

Protein-facilitated folding of the b13 group I intron ribozyme illustrates the ability of SHAPE to visualize large-scale, protein-induced conformational changes in RNA.<sup>21,49</sup> To achieve its enzymatically active splicing conformation, the b13 intron binds one copy of a maturase protein and two copies of the Mrs1 protein.<sup>50,51</sup> SHAPE analysis of the free b13 RNA revealed that much of the RNA forms base pairs that

are not consistent with the active conformation (see green-black and yellow-gray helices, Figure 8C).<sup>52</sup> The complex alternative secondary for the free RNA could be specifically tested by the SHAPE-analysis-of-point-mutations approach,<sup>42,52</sup> which readily detects long-range RNA interactions mediated by structures separated by hundreds of nucleotides. SHAPE and hydroxyl radical probing were then used to identify the binding sites for the b13 maturase and Mrs1 proteins and to determine the complex RNA structural rearrangements induced by each protein.<sup>21,49</sup>

Strikingly, protein binding induced large-scale structural rearrangements in the RNA that ultimately stabilized specific long-distance tertiary interactions (Figure 8C, red and blue arrows). Protein-mediated stabilization of tertiary interactions appears to pull the secondary structure into an energetically disfavored but functional conformation (Figure 8D).<sup>49,53</sup>

## Experimental Frontiers in RNA Structure Analysis

It is likely that RNA structure governs numerous elements of gene expression. However, until recently, it was not possible to obtain an experimentally based understanding of the relationship between RNA structure and gene function on a global scale. This general problem is by no means solved, but a number of groups have begun to explore hypotheses based on structural information collected on RNA genomes or transcriptomes.<sup>4,54,55</sup> SHAPE was recently used to probe the structure of 99% of the 9173 nucleotides in the genome of the NL4-3 strain of HIV-1,<sup>4</sup> providing the first data-guided model for the structure of an entire RNA genome. When data collected on the HIV-1 genome were smoothed over a 75-nucleotide window, we observed that regions known to be functionally important, including elements in the 5' region of the genome, at the frameshift element that allows translation of Pol, and at the Rev responsive element (RRE), tended to have low SHAPE reactivities and were thus highly structured.<sup>4</sup> The SHAPE analysis also showed that there were numerous additional highly structured regions that had not previously been associated with a biological role.

Intriguingly, there is a strong correlation between RNA structure and the domain structure of the encoded protein.<sup>4</sup> HIV-1 genome regions with low SHAPE reactivities correspond to loops between domains in the encoded protein (Figure 9). This observation is consistent with the hypothesis that RNA structure regulates the rate of ribosome processivity to facilitate protein folding. This model is supported by

ribosome mapping experiments that show that pausing occurs preferentially in regions of high levels of RNA structure<sup>4</sup> and with a study that found that yeast mRNAs tend to have a greater degree of RNA structure within their coding sequences than in their untranslated regions.<sup>54</sup>

Genome-wide SHAPE data was also used to generate a working model for the secondary structure of an entire authentic HIV-1 genome. This model includes proposals for new regulatory motifs in the genome, including the ribosome frameshifting element (Figure 6B), for structures that comprise protein domain junctions, for a domain that encodes the signal peptide in the Env protein, and for regions that have a conserved *lack* of structure, including both hypervariable sequences in Env and splice site acceptors.<sup>4</sup> Subsequent independent work has found that there is a strong correlation between HIV-1 RNA genome structure and recombination.<sup>56</sup>

## Perspective

The first analysis of an entire RNA genome revealed that the HIV-1 genomic RNA is densely populated with highly structured regions, many of which have clearly identifiable functional roles in HIV-1 replication.<sup>4</sup> Moreover, it has become possible to predict RNA structures on the kilobase scale with good accuracy.<sup>27</sup> The extent of structure in intact RNAs appears to be pervasive, to correlate extensively with key elements of gene expression, and to plausibly constitute a second level of the genetic code. The available information suggests that we have just scratched the surface regarding how biological information is encoded in the structure of RNA.

The ability of SHAPE to probe the local nucleotide environment of RNA inside viral membranes<sup>15,38</sup> opens the door to new and exciting experiments. There appear to be no significant limitations to using SHAPE in even more complex environments, including cells. RNA modification by SHAPE reagents leaves stable adducts that capture an imprint of the RNA structure during the short time the reagents are active. Subsequent purification of the RNA does not obscure this imprint and SHAPE is likely to prove broadly useful for exploring numerous structural states of RNA *in vivo*.

## BIOGRAPHICAL INFORMATION

**Kevin M. Weeks** grew up in Nashville, Tennessee, where he caught the teaching bug during many years working as a camp counselor and as a tour guide at the Country Music Hall of Fame. He began his research career and earned a B.A. at the College of Wooster in Wooster, Ohio. He spent a year in Göttingen, Germany

as a Fulbright Scholar, earned his Ph.D. with Donald Crothers at Yale, and did postdoctoral studies with Thomas Cech at the University of Colorado. Dr. Weeks is currently Professor of Chemistry at the University of North Carolina. His laboratory focuses on creating new chemical microscopes for understanding RNA structure and on applying these technologies to challenging problems in biology. The National Science Foundation and National Institutes of Health support work in his laboratory.

**David M. Mauger** was raised in Pottstown, Pennsylvania, where he first developed an interest in science. He earned a B.S. at the University of Pittsburgh, majoring in both chemistry and biology, and conducted research in yeast genetics in the laboratory of Karen Arndt. He earned his Ph.D. with Mariano Garcia-Blanco at Duke University where he studied regulation of alternative splicing in mammalian mRNAs. Dr. Mauger is currently an American Cancer Society Postdoctoral Fellow in the Weeks laboratory at the University of North Carolina at Chapel Hill. His research focuses on the development of new experimental approaches to analyze the role of RNA structure in viral replication.

## FOOTNOTES

\*To whom correspondence should be addressed. E-mail: weeks@unc.edu

## REFERENCES

- Muesing, M. A.; Smith, D. H.; Capon, D. J. Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. *Cell* **1987**, *48*, 691–701.
- Wang, Z.; Burge, C. B. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* **2008**, *14*, 802–813.
- Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **2005**, *361*, 13–37.
- Watts, J. M.; Dang, K. K.; Gorelick, R. J.; Leonard, C. W.; Bess, J. W., Jr.; Swanson, R.; Burch, C. L.; Weeks, K. M. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **2009**, *460*, 711–716.
- Komar, A. A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* **2009**, *34*, 16–24.
- Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **2005**, *127*, 4223–4231.
- Chamberlin, S. I.; Weeks, K. M. Mapping local nucleotide flexibility by selective acylation of 2'-amine substituted RNA. *J. Am. Chem. Soc.* **2000**, *122*, 216–224.
- John, D. M.; Merino, E. J.; Weeks, K. M. Mechanics of DNA flexibility visualized by selective 2'-amine acylation at nucleotide bulges. *J. Mol. Biol.* **2004**, *337*, 611–619.
- Chamberlin, S. I.; Merino, E. J.; Weeks, K. M. Catalysis of amide synthesis by RNA phosphodiester and hydroxyl groups. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14688–14693.
- Chamberlin, S. I.; Weeks, K. M. Differential helix stabilities and sites pre-organized for tertiary interactions revealed by monitoring local nucleotide flexibility in the bI5 group I intron RNA. *Biochemistry* **2003**, *42*, 901–909.
- Mortimer, S. A.; Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **2007**, *129*, 4144–4145.
- Mortimer, S. A.; Weeks, K. M. Time-resolved RNA SHAPE chemistry. *J. Am. Chem. Soc.* **2008**, *130*, 16178–16180.
- Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **2006**, *1*, 1610–1616.
- Vasa, S. M.; Guex, N.; Wilkinson, K. A.; Weeks, K. M.; Giddings, M. C. ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **2008**, *14*, 1979–1990.
- Wilkinson, K. A.; Gorelick, R. J.; Vasa, S. M.; Guex, N.; Rein, A.; Mathews, D. H.; Giddings, M. C.; Weeks, K. M. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* **2008**, *6*, No. e96.
- Steen, K. A.; Malhotra, A.; Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease. *J. Am. Chem. Soc.* **2010**, *132*, 9940–9943.

- 17 Wilkinson, K. A.; Vasa, S. M.; Deigan, K. E.; Mortimer, S. A.; Giddings, M. C.; Weeks, K. M. Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA* **2009**, *15*, 1314–1321.
- 18 Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
- 19 Shajani, Z.; Varani, G. <sup>13</sup>C NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J. Mol. Biol.* **2005**, *349*, 699–715.
- 20 Gherghe, C. M.; Shajani, Z.; Wilkinson, K. A.; Varani, G.; Weeks, K. M. Strong correlation between SHAPE chemistry and the generalized NMR order parameter ( $S^2$ ) in RNA. *J. Am. Chem. Soc.* **2008**, *130*, 12244–12245.
- 21 Duncan, C. D.; Weeks, K. M. The Mrs1 splicing factor binds the bl3 group I intron at each of two tetraloop-receptor motifs. *PLoS One* **2010**, *5*, No. e8983.
- 22 Mortimer, S. A.; Johnson, J. S.; Weeks, K. M. Quantitative analysis of RNA solvent accessibility by N-silylation of guanosine. *Biochemistry* **2009**, *48*, 2109–2114.
- 23 Mathews, D. H.; Turner, D. H. Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* **2006**, *16*, 270–278.
- 24 Low, J. T.; Weeks, K. M. SHAPE-directed RNA secondary structure prediction. *Methods* **2010**, *52*, 150–158.
- 25 Dowell, R. D.; Eddy, S. R. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinf.* **2004**, *5*, No. 71.
- 26 Mathews, D. H. RNA secondary structure analysis using RNAstructure. In *Current Protocols in Bioinformatics*; Baxevanis, A. D., Davidson, D. B., Page, R. D., Petsko, G. A., Stein, L. D., Stormo, G. D., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2006; pp 12.6.1–12.6.14.
- 27 Deigan, K. E.; Li, T. W.; Mathews, D. H.; Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 97–102.
- 28 Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7287–7292.
- 29 Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **1999**, *288*, 911–940.
- 30 Mauger, D. M.; Weeks, K. M. Toward global RNA structure analysis. *Nat. Biotechnol.* **2010**, *28*, 1178–1179.
- 31 Jacks, T.; Power, M. D.; Masiarz, F. R.; Luciw, P. A.; Barr, P. J.; Varmus, H. E. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **1988**, *331*, 280–283.
- 32 Dulude, D.; Baril, M.; Brakier-Gingras, L. Characterization of the frameshift stimulatory signal controlling a programmed –1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res.* **2002**, *30*, 5094–5102.
- 33 Dinman, J. D.; Richter, S.; Plant, E. P.; Taylor, R. C.; Hammell, A. B.; Rana, T. M. The frameshift signal of HIV-1 involves a potential intramolecular triplex RNA structure. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5331–5336.
- 34 D'Souza, V.; Summers, M. F. How retroviruses select their genomes. *Nat. Rev. Microbiol.* **2005**, *3*, 643–655.
- 35 Badorrek, C. S.; Weeks, K. M. Architecture of a gamma retroviral genomic RNA dimer. *Biochemistry* **2006**, *45*, 12664–12672.
- 36 Gherghe, C.; Weeks, K. M. The SL1-SL2 (stem-loop) domain is the primary determinant for stability of the gamma retroviral genomic RNA dimer. *J. Biol. Chem.* **2006**, *281*, 37952–37961.
- 37 Gherghe, C.; Leonard, C. W.; Gorelick, R. J.; Weeks, K. M. Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA dimerization domain. *J. Virol.* **2010**, *84*, 898–906.
- 38 Gherghe, C.; Lombo, T.; Leonard, C. W.; Datta, S. A.; Bess, J. W., Jr.; Gorelick, R. J.; Rein, A.; Weeks, K. M. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19248–19253.
- 39 Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J. Am. Chem. Soc.* **2005**, *127*, 4659–4667.
- 40 Wang, B.; Wilkinson, K. A.; Weeks, K. M. Complex ligand-induced conformational changes in tRNA(Asp) revealed by single-nucleotide resolution SHAPE chemistry. *Biochemistry* **2008**, *47*, 3454–3461.
- 41 D'Souza, V.; Summers, M. F. Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* **2004**, *431*, 586–590.
- 42 Badorrek, C. S.; Weeks, K. M. RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat. Chem. Biol.* **2005**, *1*, 104–111.
- 43 Gherghe, C. M.; Mortimer, S. A.; Krahn, J. M.; Thompson, N. L.; Weeks, K. M. Slow conformational dynamics at C2'-endo nucleotides in RNA. *J. Am. Chem. Soc.* **2008**, *130*, 8884–8885.
- 44 Mortimer, S. A.; Weeks, K. M. C2'-endo nucleotides as molecular timers suggested by the folding of an RNA domain. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15622–15627.
- 45 Julien, K. R.; Sumita, M.; Chen, P. H.; Laird-Offringa, I. A.; Hoogstraten, C. G. Conformationally restricted nucleotides as a probe of structure-function relationships in RNA. *RNA* **2008**, *14*, 1632–1643.
- 46 Krasilnikov, A. S.; Yang, X.; Pan, T.; Mondragon, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **2003**, *421*, 760–764.
- 47 Loria, A.; Pan, T. Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry* **1997**, *36*, 6317–6325.
- 48 Mortimer, S. A.; Weeks, K. M. Time-resolved RNA SHAPE chemistry: Quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.* **2009**, *4*, 1413–1421.
- 49 Duncan, C. D.; Weeks, K. M. Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding. *Biochemistry* **2010**, *49*, 5418–5425.
- 50 Bassi, G. S.; de Oliveira, D. M.; White, M. F.; Weeks, K. M. Recruitment of intron-encoded and co-opted proteins in splicing of the bl3 group I intron RNA. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 128–133.
- 51 Bassi, G. S.; Weeks, K. M. Kinetic and thermodynamic framework for assembly of the six-component bl3 group I intron ribonucleoprotein catalyst. *Biochemistry* **2003**, *42*, 9980–9988.
- 52 Duncan, C. D.; Weeks, K. M. SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* **2008**, *47*, 8504–8513.
- 53 McGinnis, J. L.; Duncan, C. D.; Weeks, K. M. High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol.* **2009**, *468*, 67–89.
- 54 Kertesz, M.; Wan, Y.; Mazor, E.; Rinn, J. L.; Nutter, R. C.; Chang, H. Y.; Segal, E. Genome-wide measurement of RNA secondary structure in yeast. *Nature* **2010**, *467*, 103–107.
- 55 Underwood, J. G.; Uzilov, A. V.; Katzman, S.; Onodera, C. S.; Mainzer, J. E.; Mathews, D. H.; Lowe, T. M.; Salama, S. R.; Haussler, D. FragSeq: Transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **2010**, *7*, 995–1001.
- 56 Simon-Loriere, E.; Martin, D. P.; Weeks, K. M.; Negroni, M. RNA structures facilitate recombination-mediated gene swapping in HIV-1. *J. Virol.* **2010**, *84*, 12675–12682.